Supplementary Note, Supplementary Figures 1–8, and Supplementary Tables 1, 2, 6, 7, 11, 12, 15, 16, 20, 24, and 25 for:

Genomics of *Loa loa*, a *Wolbachia*-free filarial parasite of humans

Christopher A. Desjardins[1], Gustavo C. Cerqueira[1], Jonathan M. Goldberg[1], Julie C. Dunning Hotopp[2], Brian J. Haas[1], Jeremy Zucker[1], Jose' M.C. Ribeiro[4], Sakina Saif[1], Joshua Z. Levin[1], Lin Fan[1], Qiandong Zeng[1], Carsten Russ[1], Jennifer R. Wortman[1], Doran L. Fink[3]*, Bruce W. Birren[1], Thomas B. Nutman[3]

[1]Broad Institute of MIT and Harvard, Cambridge, MA

[2]Institute for Genome Science, Department of Microbiology & Immunology, University of Maryland School of Medicine, Baltimore, MD

[3]Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, MD

[4]Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, Bethesda, MD

*current address: Center for Biologics Evaluation and Research, Food and Drug Administration, Rockville, MD

Supplementary Note

Repeat analysis

The most prominent repeats in the *L. loa* genome are the BEL retrotransposons that together comprised 1.3% of the genome, with the single most abundant repeat BEL-3 comprising 0.62% of the genome. The previously identified *L. loa* repeat LL3M9[1] used for molecular diagnostics is the sixth most abundant repeat in the genome totaling an estimated 212 copies and 0.20% of the genome. Like *L. loa*, the *W. bancrofti* genome has BEL repeats that comprised 1.54% of the genome; the SSP1[2] repeat, used for molecular diagnosis of *W. bancrofti,* was estimated to contain 2,142 copies comprising 0.51% of the genome (see Methods).

Annotation

Of the 17.6 million RNA-Seq reads, approximately 65% had high quality paired alignments to the *L. loa* genome. Annotation improvements based on RNA-Seq data include the identification of 502 alternatively spliced genes and the addition of 5128 and 5238 5' and 3' UTRs, respectively (Supplementary Table 16). Expression levels of each gene based on these alignments is shown in Supplementary Table 17. CEGMA[3] analysis found that 236/248 (95%) core eukaryotic genes are complete in the *L. loa* assembly, while an additional 5 are partial. In the *W. bancrofti* assembly, 165/248 (67%) are complete and an additional 51 are partial.

Fragmented genes in draft assemblies

Due to the high degree of fragmentation of the *W. bancrofti* genome assembly when compared to the *B. malayi* and *L. loa* assemblies (Table 1), it is likely that the predicted *W. bancrofti* gene count is an over-estimation of the total gene content; in some cases predicted genes probably represent pieces of a single actual gene that could not be assembled together into a contiguous sequence. In support of this hypothesis, 88% of *L. loa* genes are full length while only 59% of *W. bancrofti* genes are full length. Based on comparisons to orthologs in *L .loa* and *B. malayi*, the true gene count of *W. bancrofti* is estimated to be 14,496–15,075, in the range of the *L. loa* gene count. The *L. loa* genome is much less fragmented; based on comparison to orthologs in *B. malayi*, the true gene count of *L. loa* is estimated to be 14,261, within 4.5% of the original count of 14,925.

Horizontal transfer of viral RNA helicases

An RNA helicase domain, involved in viral DNA replication, was found to be enriched in the *L. loa* genome relative to other filarial nematodes ($p < 0.05$, Fisher's exact test). All proteins with this domain were intronless and in some cases the proteins were tandemly repeated. In no case was another gene with similarity to viral genes found adjacent to these *L. loa* genes. These proteins were most similar to those belonging to cycloviruses (BLAST p-value $< 1e^{-10}$), a recently described group

of ssDNA viruses with a broad animal host range[4]. It is possible that these genes were horizontally transferred into *L. loa* following infection by this virus. Examination of RNA-Seq data revealed that two of these genes have expression levels in the top 3% of all genes in microfilariae.

Immunologically related gene products

The TLR related pathway molecules include: a putative interleukin-1-receptor-associated kinase, toll-like receptor adaptors, TNF-receptor-associated factor 4 (TRAF4) and I Kappa Kappa (iKK). Thus, *L. loa* (like the other filarial nematodes) appear to possess a primordial Toll-related pathway as an early line of defense against microbial infections.

The potential *L. loa* encoded autoantigens have been implicated in autoimmune processes -- systemic lupus erythematosus (KU70/80), type 1 diabetes mellitus (GAD, ida-1), Sjoegren's syndrome and systemic sclerosis (Sjoegren's syndrome/scleroderma autoantigen 1 homolog, Major centromere autoantigen B, Major centromere autoantigen B, crn-3), primary biliary cirrhosis (Nuclear autoantigen Sp-100) -- as well as those that are more non-specific -- uveal autoantigen, 2 golgi autoantigens, and NGP-1 autoantigen.

Protein kinase annotation

Annotations of the complete kinomes of *L. loa*, *W. bancrofti* and other nematodes used for comparison in this study, using the *C. elegans* kinome (www.kinase.com) as a reference, are detailed in Supplementary Tables 21 and 22. All major protein kinase groups are present in all of the nematodes analyzed. Based on these methods and utilizing a recent release of the *B. malayi* annotation (release 230 from www.wormbase.org), we identified 67 kinases in addition to those previously described for *B. malayi*[5]. Of the 310 protein kinases identified in *L. loa*, 250 have orthologs in *C. elegans* (Supplementary Tables 20 and 22). This means that in spite of its smaller kinome, *L. loa* has 60 protein kinases without orthologs in *C. elegans*. These kinases are conserved in the other filarial worms suggesting they serve important functions in filarial biology. Eight protein kinases from known families present in *L. loa* are missing from *C. elegans* (Supplementary Table 22); these widely-conserved kinases were likely present in ancestral nematodes and lost in Rhabdtina. The best characterized kinases from this set include Fray, involved in axon-sheathing[6], SLOB, an inactive kinase-like molecule that allosterically regulates the slowpoke potassium channel[7], CDK10, which activates Ets2-pointed transcription factors in a phosphorylation-independent manner[8], TESK, a cytoskeleton regulator that acts through ADF/cofilin phosphorylation[9], and TTK (see below and main text). Six *L. loa* protein kinases have no known family affiliation but are conserved with other nematodes, as shown in a phylogenetic analysis of unclassified nematode protein kinases (Supplementary Fig. 5).

L. loa lacks orthologs of 160 *C. elegans* kinases (Supplementary Tables 20 and 22). These losses occur in other filarial worms as well, suggesting that they are not defects in assembly or annotation. Most of these losses occur in one of a few large expansions in the non-parasitic nematodes, including CK1/TTBKL (13 fewer kinases), CK1/Worm6 (14 fewer kinases), Haspin (19 fewer kinases), RGC (18 fewer kinases) and TK/FER (22 fewer kinases). The best characterized *C. elegans* expansion is in the RGC (receptor guanylate cyclase) group, which includes kinases involved in environmental sensing; for example, *gcy-22* is required in the process of worms learning to associate the presence of NaCl with the presence or absence of food[10], and worms with *gcy-14* mutations are defective in Na+ and Li+ chemotaxis[11]. There are also kinases from nine families present in *C. elegans* that have neither an ortholog nor a non-orthologous relative in *L. loa* (Supplementary Table 22). These kinases include the non-RGC sensory kinase *sgk-1*, which is involved in the integrative response to an olfactory diacetyl and a gustatory Cu (2+) stimuli[12], C01C4.3 and *kin-22* kinases, which are involved in brain and eye development in zebrafish and fly, respectively[13,14], and *kin-29*[15], which regulates the expression of the neuronal chemoreceptor gene *str-1* in response to environmental factors. The target of *kin-29*, *str-1*, is also missing from the filarial worms (see main text). The absence of these kinases and this substrate from filarial worms reinforces the hypothesis that these parasitic worms inhabit an environment that is less complex in terms of olfactory and gustatory information than that inhabited by soil nematodes.

An additional family loss in *L .loa* and other filarial worms with respect to Rhabdtina is that of the nearly universally conserved RAD53-family kinase *chk-2* (see main text). In most eukaryotes RAD53 plays a role in initiating cell-cycle arrest when DNA damage is present[16]. It does not function in this way in *C. elegans*; rather it has a role in the regulation of meiosis that is not conserved in other species. As noted above TTK, a widely conserved kinase that regulates exit from meiosis[17], is missing from *C. elegans* but present in filarial worms. The reciprocal distribution of RAD53 and TTK suggests that in filarial worms meiosis is regulated in a manner more similar to other metazoans. The phylogenetic distribution of RAD53 and TTK is complicated by the fact that RAD53 and TTK coexist in *A. suum* and *P. pacificus* (Supplementary Table 22). This suggests that the mechanisms regulating meiosis are present in *P. pacificus* and *A. suum*, and were probably present in the common ancestor of the Rhabdtina and Spiruria. The number of protein kinases detected in *T. spiralis* and *M. hapla* is sharply lower than for the other nematodes (Supplementary Table 22), which may be an artificial result caused by incomplete genome coverage for these species. Therefore, the absence of RAD53 and TTK from *T. spiralis* and *M. hapla* is uncertain.

Metabolic profiling of nematode and *Wolbachia* genomes

As heme, riboflavin, FAD, glutathione and nucleotide biosynthetic pathways have been previously implicated in the filaria-*Wolbachia* symbiosis[18], we examined the distribution of the relevant metabolic pathways across sequenced nematode and

Wolbachia genomes (Tables 2, Supplementary Tables 24, 25). No differences in these pathways exist between *L. loa* and other filarial nematodes, or between nematode and insect *Wolbachia*.

Within the heme biosynthesis pathway, ferrochelatase, which catalyzes the final step in heme synthesis and is found in the *B. malayi* genome, was also found in the filarial nematodes *O. volvulus*, *D. immitis*, and the *Wolbachia*-free animal parasite *A. vitae*, and is hypothesized to have been laterally transferred from a gamma proteobacterial ancestor unrelated to *Wolbachia*[19]. It had also been noted that *wBm* is missing a single member of the heme biosynthesis pathway, hemG/protoporphyrinogen oxidase, a gene which had not been identified in a number of gram-negative bacteria[18]. It has recently been shown that COG1981 can act in the same capacity as hemG[20], and we identified this gene in all analyzed *Wolbachia* genomes (Wbm0208 in *wBm*, WD0417 in *wMel*, WP0085 in *wPip*, and present in *wWb*).

The animal flavin biosynthesis pathway, which synthesizes FAD from riboflavin, was identified as complete in all filarial and most other nematode genomes (Supplementary Table 24). Riboflavin kinase was not identified in *P. pacificus* or *M. hapla* (Table 2, Supplementary Table 24). However, riboflavin kinase is a relatively small gene (135 amino acids in *C. elegans*) and it may be present but difficult to identify a divergent copy by sequence similarity, or it may be present in the genome but not covered in the draft assemblies. Three of the four examined *Wolbachia*

genomes have complete bacterial flavin biosynthesis pathways where they can synthesize riboflavin *de novo* and subsequently synthesize FAD utilizing a different pathway than animals. However, only half of the genes involved in flavin biosynthesis were identified in *wWb* genome. Although it is possible that this pathway is in the process of being lost so that only a fraction of the genes are still present, the low sequence coverage of the *wWb* genome (2X) makes it more likely that the pathway is present but those specific genes were not covered in the draft genome. Because animals cannot synthesize riboflavin *de novo*, even if *Wolbachia* were involved in riboflavin supplementation in some filarial nematodes, all nematodes likely have mechanisms for acquiring riboflavin from the environment.

The filarial genomes encode only a single gene from the purine biosynthesis pathway, adenylsuccinate lyase. This gene also functions in the purine interconversion pathway, and its necessity for this pathway may be what has caused it to be maintained in the filarial genomes.

In addition to the five previously hypothesized pathways, we profiled vitamin B6 (pyridoxine) synthesis after noting differences among filarial and *Wolbachia* genomes. This pathway involves two enzymes: pyridoxine kinase and pyridoxal 5'-phosphate synthase. Most nematode genomes encode a single copy of each enzyme. However, the *C. briggsae* genome encodes one additional copy of pyridoxal 5'-phosphate synthase while the *L. loa* genome encodes four additional copies.

Metabolic transporters in nematode and *Wolbachia* genomes

All known transporters relating to the five metabolic pathways being examined were profiled in all nematode genomes based on either PFAM domains or similarity to characterized transporters. Numerous equilibrant nucleoside transporters were identified in the nematode genomes based on PFAM domains. We therefore constructed a phylogenetic tree of all homologous genes (Supplementary Fig. 8). There are no expansions of these transporters either within *L. loa* or the *B. malayi-W. bancrofti* clade. Riboflavin transporter 2, a known transporter of riboflavin in *C. elegans*, was identified in all nematode genomes. No glutathione transporter could be profiled as the only known eukaryotic transporter, the yeast gene Hgt1p, does not have homologs in other eukaryotes including *C. elegans*[21].

We did identify a duplication of a purine-specific 5' nucleotidase in the filarial genome. While the functional significance of its divergent residues is unknown, this additional nucleotidase could provide filarial worms with additional or alternate ways to salvage needed purines. Furthermore, the purine salvage gene GMP reductase was mentioned as one of the most promising druggable targets in the *A. suum* genome[22].

We also searched for the *C. elegans* heme-responsive genes (hrg's) 1-6, known to be involved in the transport of heme[23,24]. We identified an ortholog of *hrg-1* in all nematode genomes. Genes *hrg-3*, *hrg-4*, and *hrg-5* were restricted to *Caenorhabditis*,

while *hrg-2* and *hrg-6* were also found in *P. pacificus*. It also should be noted that *hrg-3* through *-6* all share sequence similarity with *hrg-1* and may have resulted from numerous evolutionary duplication events. We found no evidence of similar duplications of the filarial nematode ortholog of *hrg-1*, so it is unclear how similar their heme transport mechanisms are to those of *C. elegans*.

We additionally attempted to identify pyroxidine transporters in both the nematode and *Wolbachia* genomes. However, we could not identify a homolog of the only characterized eukaryotic pyridoxine transporter, the yeast gene Tpn1p[25], in any of the nematode genomes. Furthermore, the proteobacterial transporter for pyridoxine is also unknown, as Firmicutes have a transporter PdxT, which is not found in other bacteria[26]. In *E. coli*, PdxT is adjacent to the other pyridoxine synthesis genes (PdxJ and PdxH), but in *wBm* there were no transporters adjacent to these genes. Given the absence of differences in relevant transporters between *L. loa* and *Wolbachia*-containing filarial parasites, it is unlikely that unique nematode transporters are needed to either maintain the symbiotic relationship between the filaria and *Wolbachia*.

Nuwts in the *L. loa* genome

An initial search of the *Wolbachia of B. malayi* genome against the *L. loa* genome done using BLASTN (cutoff = 1e-5) revealed no matches over 200 bp. We therefore turned to a read-based search strategy to identify nuclear *Wolbachia* transfers

(nuwts; see Methods). Using this approach, 15 putative nuwts were detected in the *L. loa* genome, ranging in size from 33-188 bp with an average size of 76 bp. They are 81-100% identical in nucleotide sequence to the *Wolbachia* wBm genome with an average of 86% nucleotide identity.

The putative nuwts in *L. loa* were assigned to one of three categories based on their location in the genomic regions (Supplementary Table 15). The class I putative nuwts have lateral gene transfer (LGT) of *Wolbachia* DNA in an exon of genes involved in mitochondrial energy metabolism and cannot be definitively assigned as having arisen from a *Wolbachia* endosymbiont. The transfers in classes II and III most likely arose from a *Wolbachia* endosymbiont, but do not show evidence of transcription, which would suggest functionality. The class II putative nuwts have LGT of *Wolbachia* DNA in the introns of genes encoding hypothetical proteins. The three genomic regions with class III putative nuwts have LGT of *Wolbachia* DNA in areas of the genome devoid of a known gene.

Putative *L. loa* class I nuwts may not have arisen from LGT

All of the class I putative nuwts are in all filarial nematodes genomes sequenced. For these nuwts, the putative integrations occur in numts, genes of mitochondrial ancestry that are now encoded in the nematode nucleus. Microhomology, meaning short regions of homology, between mitochondrial genes and *Wolbachia* genes might be expected since *Wolbachia* are related to mitochondria (e.g. [27]). However,

mitochondria invaded the eukaryotic lineage early in the lineage's history and as such, remnants from the mitochondria should have diverged significantly unless the regions are under significant functional constraint or the organisms have converged on the same sequence.

Therefore, we sought to investigate whether microhomology could be detected between *Wolbachia* genes and other *Wolbachia*-free eukaryotic genomes. We were not able to identify nuwts in *Nematocida parisii,* which was sequenced at the same center as *L. loa* and had similar sequencing statistics. However, *N. parisii* is a microsporidia that lacks mitochondria. Therefore, a second comparison was made to the parasitic nematode *Trichinella pseudospiralis* (SRX000172) sequenced at Washington University Genome Sequencing Center that had similar sequencing statistics. In a similar screen to *L. loa*, 23 unique reads were identified with homology to *w*Bm. While this is fewer than the 248 detected in *L. loa,* the identified reads also have significant matches to numts. Since numts were in an unrelated nematode, it raises the possibility that all putative class I nuwts are not really nuwts, but instead are highly conserved sequences in numts.

Relation of putative class II and class III nuwts with other *Wolbachia* strains and sister taxa

Putative class II and class III nuwts were more readily determined to have arisen through LGT. None of these genes are homologous to numts simplifying the analysis.

Additionally, the only nucleotide sequences with significant homology are from *Wolbachia* endosymbionts and the protein sequences have high homology only to filarial nematodes and bacteria. To determine the donor of the class II and III nuwts, maximum likelihood phylogenies were generated for putative nuwts that have a sufficient number of homologous nucleotide sequences (Supplementary Fig. 2). While closely related *Wolbachia* strains could be clustered with significant bootstrap values (>60%), none of the nuwts could be clustered with any other *Wolbachia* strain. Therefore, they may have arisen from the *Wolbachia* strain in either the nematode or its insect host, although the nematode *Wolbachia* strain is the more likely donor. The nuwts had significantly longer branch lengths than their counterparts in other *Wolbachia* strains. Consistent with that observation, their alignments displayed a larger number of unique SNPs relative to comparisons between *Wolbachia* strains. For NUWT8, a maximum likelihood analysis as implemented in RAxML was conducted on the corresponding peptide sequence using only sequences from NR. It shows the same branching as the nucleotide tree and shows that the *Loa loa* sequence is in the *Wolbachia* lineage (Supplementary Fig. 2F).

The significant divergence of filarial nematode nuwts in *Wolbachia*-free nematode lineages from all other *Wolbachia* strains has been observed previously [28]. This may indicate that nuwts in *Wolbachia*-free lineages are acquiring mutations more rapidly than their endosymbiont equivalents. Alternatively, they may have acquired such transfers from a more divergent *Wolbachia* strain. For example, supergroup F

Wolbachia strains have been reported in filarial nematodes, but sequences were not available for this comparison. Most intriguingly, this observation may indicate that such transfers happened prior to the radiation of *Wolbachia* strains, and may suggest that the *Wolbachia* lineage arose in nematodes as a mutualistic symbiont and later was transferred to insects where it has been exceedingly successful as a parasitic symbiont.
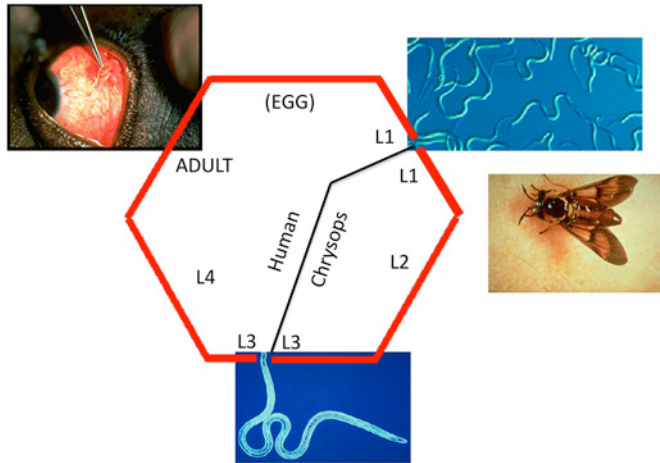
Supplementary Figure 1. *Loa loa* has a two phase life cycle encompassing the human definitive host and a deerfly *(Chrysops)* vector host. In the human host, adult males and females reside in the subcutaneous tissue and viviparously release microfilariae (~270 um in length) into the bloodstream. Microfilariae are developmentally arrested until they are taken up in a blood meal by the female deerfly. In the fly, they resume development, undergo 2 molts and migrate to the mouthparts of the deerfly as third stage larvae (L3) where they are again in an arrested state. They are then introduced into the human host during the next blood feeding episode. The L3 resume growth and development, molt twice to become motile adults that migrate through the subcutaneous tissue. Microfilarial release occurs after ~ 120 days[19], and the 40-70 mm long adult females can live for 15-20 years. In the image, the two stages (L1 and L3) that are found in both the vector and the human are depicted as is an adult worm at the time of removal from the subconjunctival space of a *Loa*-infected individual. The *Chrysops* vector is also shown.
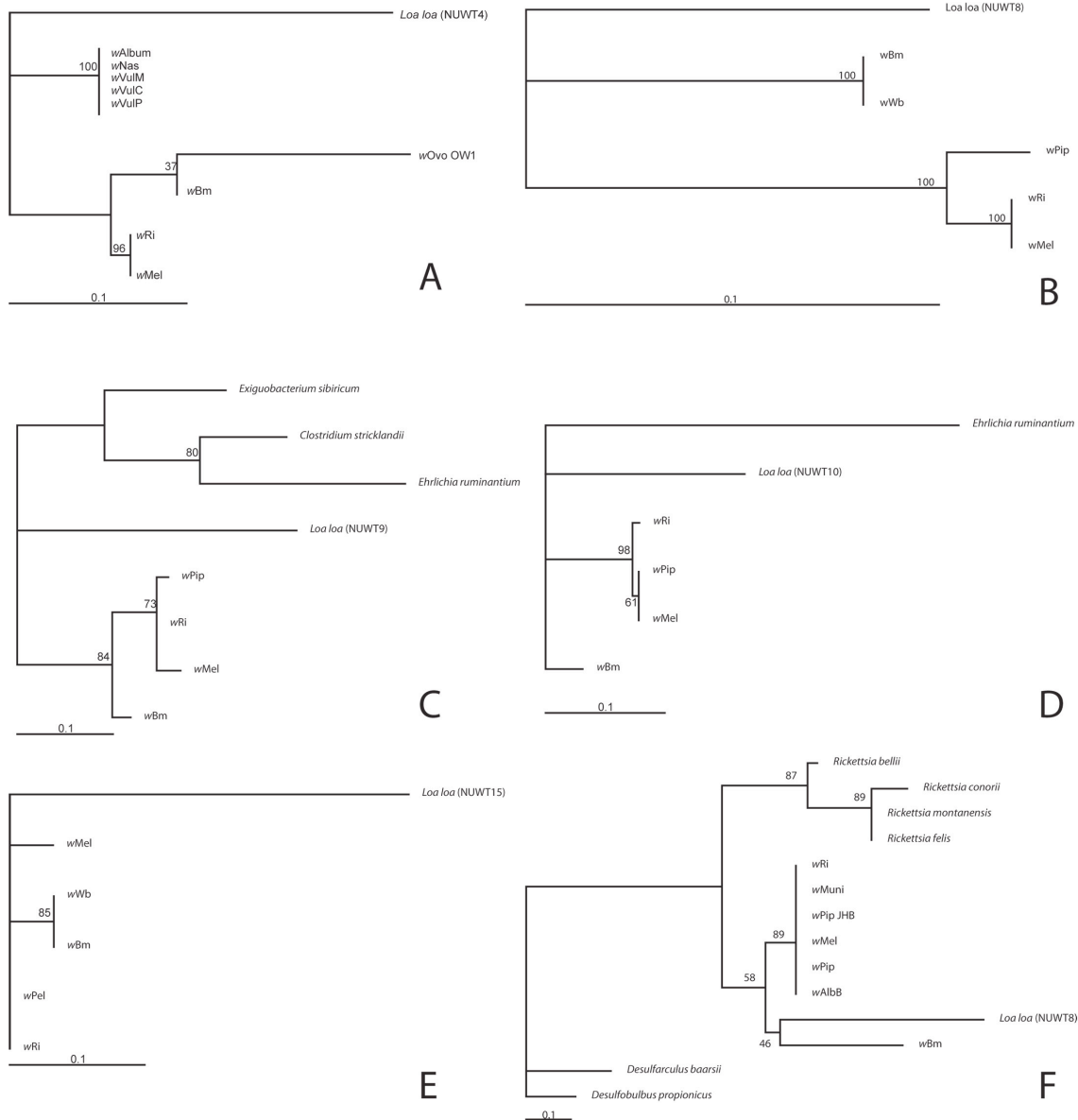
Supplementary Figure 2. Phylogenetic relationships for select class II and class III putative nuwts. Three putative class II nuwts and two putative class III nuwts were examined using maximum likelihood as implemented in RAxML ((A) NUWT4, (B) NUWT8, (C) NUWT9, (D) NUWT10, and (E) NUWT15). In all cases, phylogenetic analysis supports that these sequences arose from LGT from a *Wolbachia* strain, but the precise donor could not be determined. The nuwts are significantly diverged from their *Wolbachia* strains with longer branch lengths compared to their counterparts. For one such putative transfer ((F) NUWT8), a maximum likelihood analysis as implemented in RAxML was conducted on the corresponding peptide sequence using only sequences from Genbank's NR database. This analysis produced the same branching as the nucleotide tree and shows that the *L. loa* sequence is firmly rooted in the bacterial tree.

A

B. malayi	W. bancrofti	A. suum	C. elegans	
5889	1907	1802	304	*L. loa*
	1573	1152	229	*B. malayi*
		349	80	*W. bancrofti*
			447	*A. suum*

B
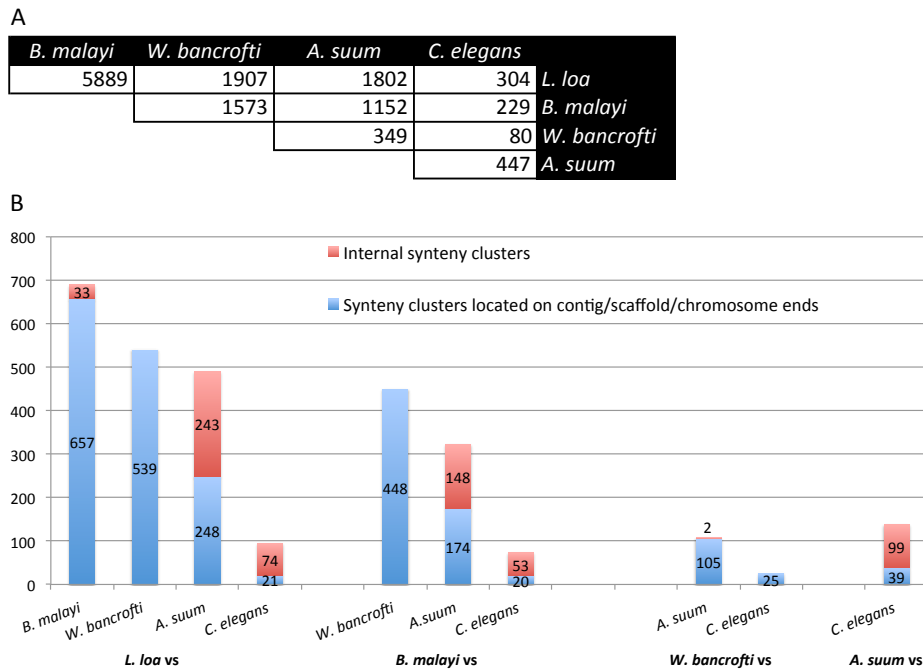
Supplementary Figure 3. A) Total number of syntenic genes based on pairwise comparisons between *L. loa*, *B. malayi*, *W. bancrofti*, *A. suum*, and *C. elegans*. Only synteny clusters with more than 3 genes were utilized. **B)** Number of synteny clusters identified on each pairwise comparison between *L. loa*, *B. malayi*, *W. bancrofti*, *A. suum*, and *C. elegans*. An entire bar indicates the total number of clusters identified in each analysis. The total number of clusters was then divided into 1) clusters located internally on scaffolds or chromosomes (red bar), and 2) those near chromosomes or scaffold ends (blue bar).

A)

B)
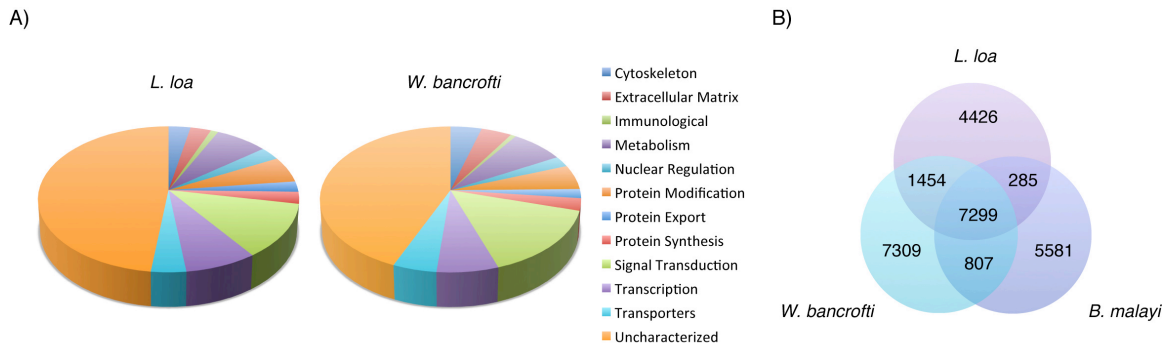
Supplementary Figure 4. Overview of function and orthology in filarial genomes. A) Functional annotation of the *L. loa* and *W. bancrofti* transcriptomes. Each pie slice represents the percentage of each functional category as a proportion of the total transcriptome. Only a single category was assigned to each gene. B) Venn diagram of shared ortholog clusters and unique gene content among the filarial genomes.
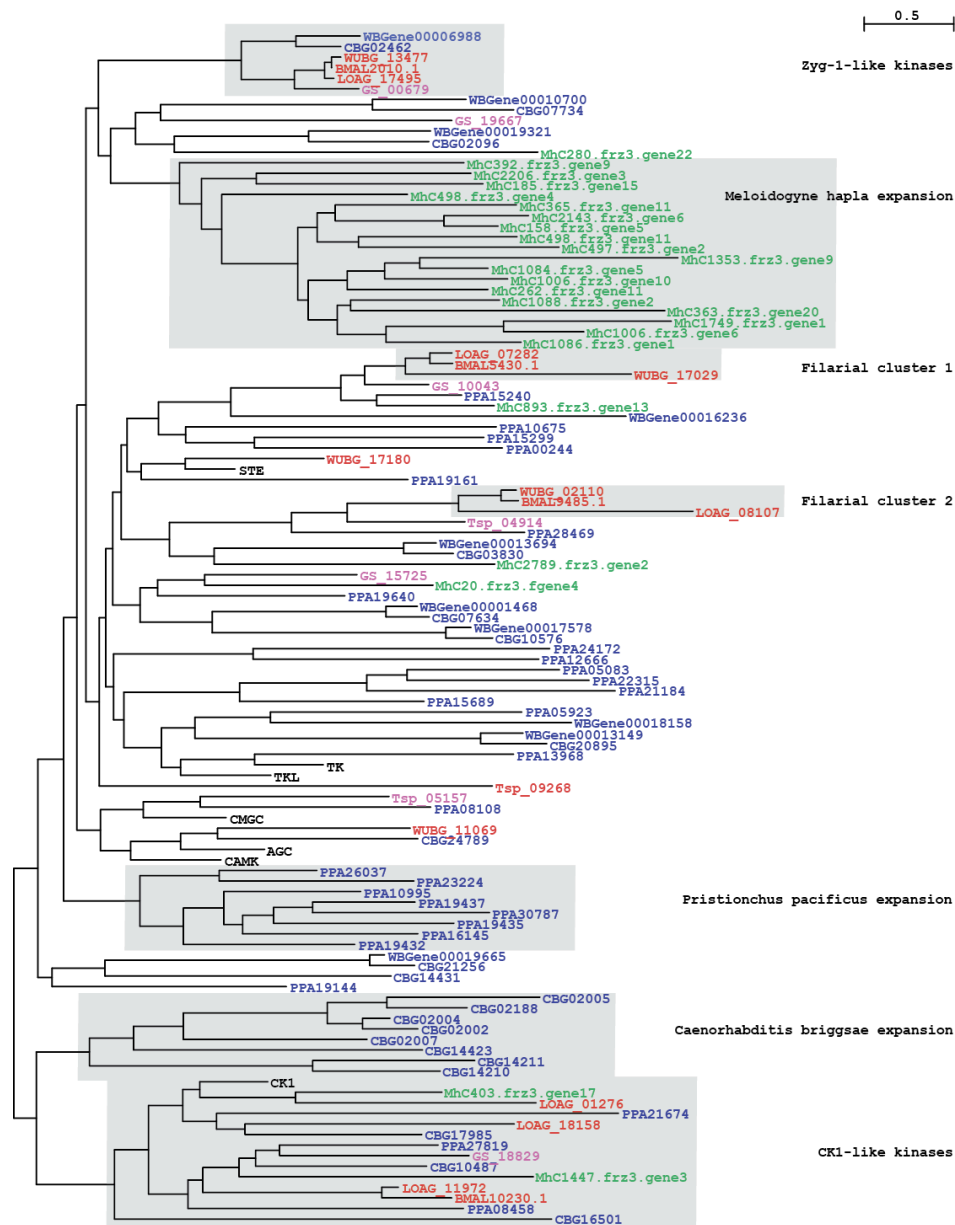
Supplementary Figure 5. Phylogenetic relationships of unclassified, potentially novel, nematode protein kinases. The sequences of unclassified kinases from nine nematodes were aligned using hmmer 3[29], and a tree based on this alignment was obtained using FastTree[30]. Gene names are colored as follows: red, filarial nematodes; blue, non-parasitic species; purple, *A. suum*; green, *M. hapla*; pink, *T. spiralis*. *C. elegans* identifiers are from Wormbase (www.wormbase.org). Consensus sequences from the major kinase groups are included for reference.
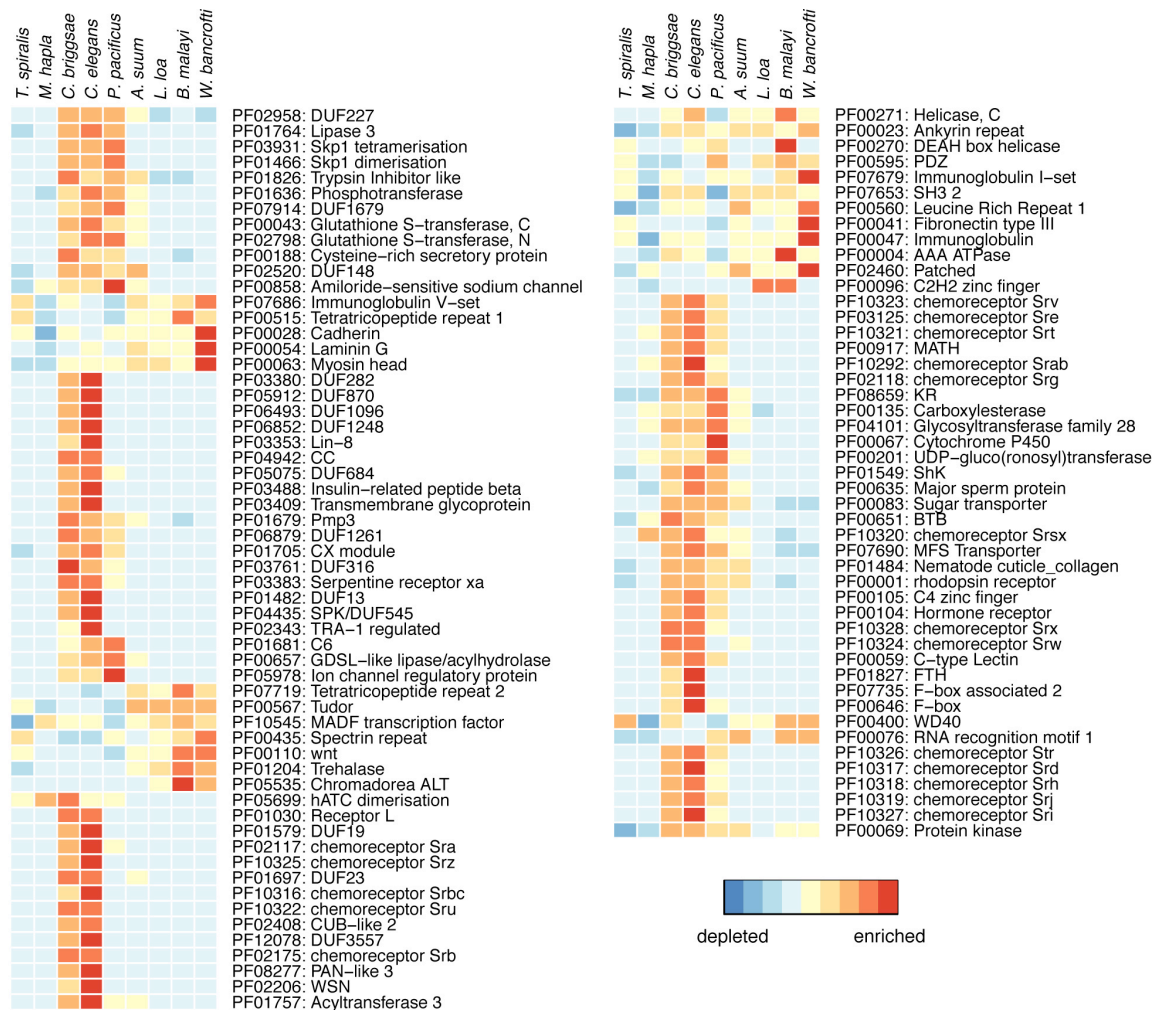
Supplementary Figure 6. PFAM domains enriched in either filarial or non-parasitic nematodes. All PFAM domains that differed in total count between the filarial (*L. loa*, *W. bancrofti*, and *B. malayi*) and free-living (*C. elegans*, *C. briggsae*, and *P. pacificus*) nematodes were compared with Fisher's exact test and subsequently corrected for multiple comparisons using FDR[31]. Only domains with a q-value < 0.05 are shown. Seven transmembrane G-protein-coupled chemoreceptor (7 TM GPCR) domains are labeled as chemoreceptors.
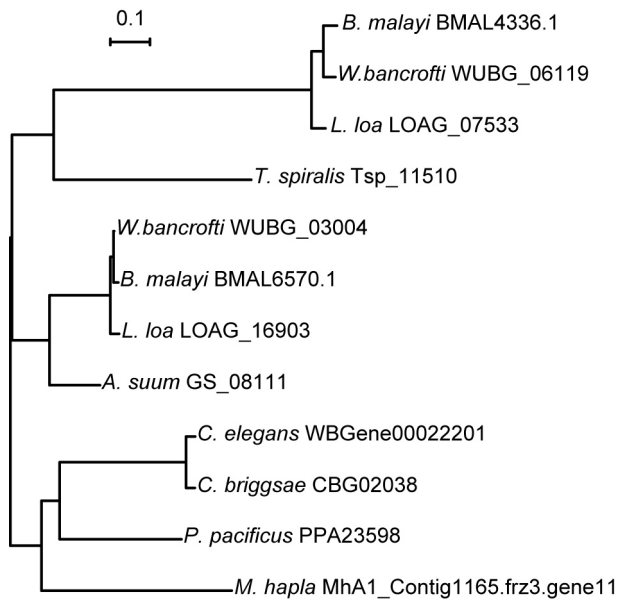
Supplementary Figure 7. Phylogenetic relationships of purine-specific 5'
nucleotidases in nematodes. *C. elegans* identifiers are from Wormbase
(www.wormbase.org). Amino acid alignments were built using Muscle[32], followed
by maximum likelihood phylogenetic reconstruction as implemented in FastTree[30].
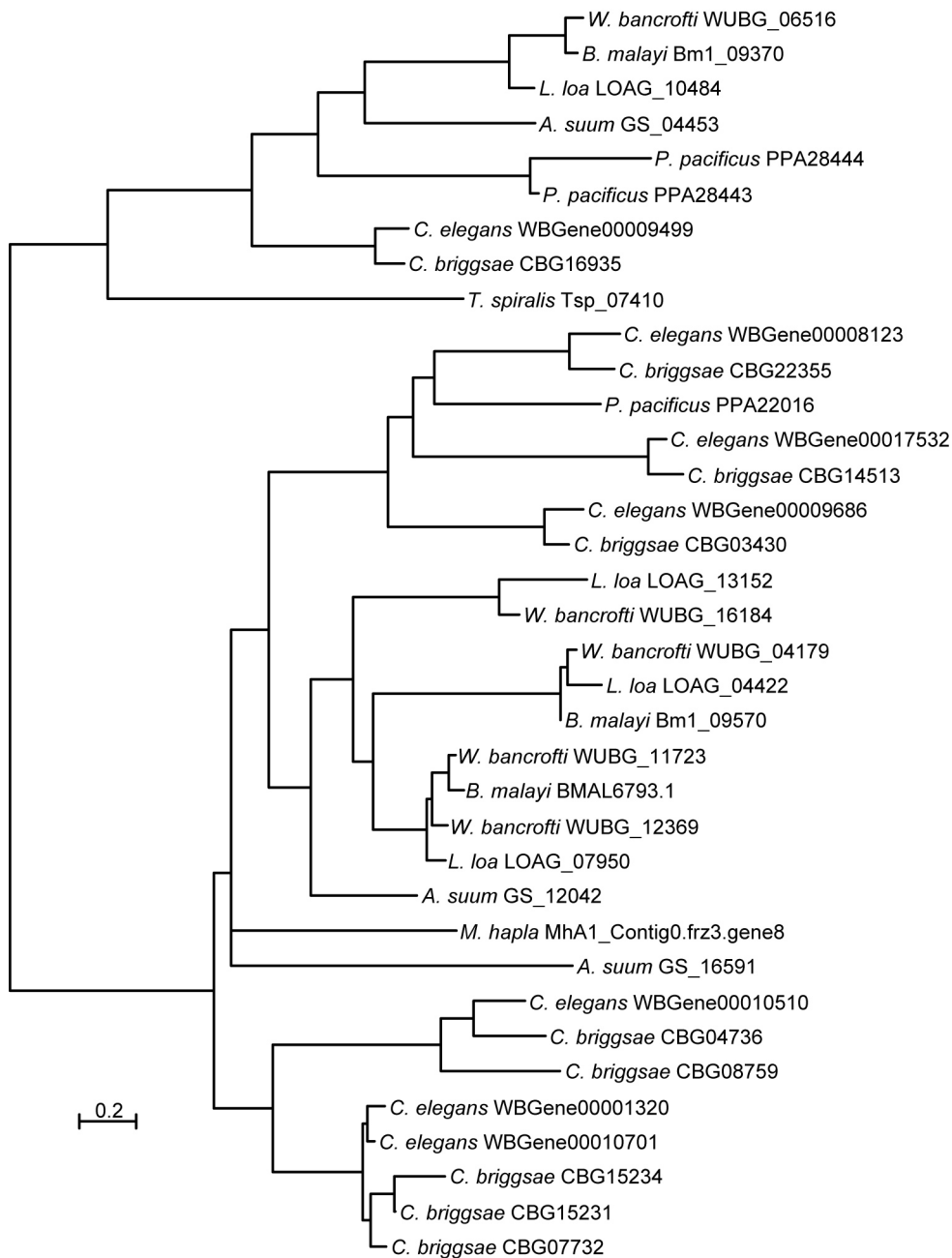
Supplementary Figure 8. Phylogenetic relationships of equilibrant nucleoside transporters in nematodes. *C. elegans* identifiers are from Wormbase (www.wormbase.org). Amino acid alignments were built using Muscle[32], followed by maximum likelihood phylogenetic reconstruction as implemented in FastTree[30].

Supplementary Table 1. Summary of repetitive and low-complexity sequence in the *L. loa* genome.

Category	Overall length (kb)	Adjusted length (kb)	Fold difference between overall length and adjusted length	Percent of genome	Percent of genome (adjusted)
Repetitive elements	3524	8455	2.4	3.88	9.31
Low complexity regions	1634	1553	0.95	1.80	1.71

Supplementary Table 2. Known nematode repeats found in the *L. loa* genome.

Repeat	Overall length (kb)	Adjusted length (kb)	Fold difference between overall length and adjusted length	Percent of genome	Percent of genome (adjusted)	Repeat unit length	Estimated #copies	Estimated #copies (adjusted length)
LLRP1	3.5	287.40	83.11	0.0038	0.3163	817	4.2	352
LL3M9[1]	2.92	177.60	60.84	0.0032	0.1955	839	3.5	212
BEL-3_BMa-I	304.5	561.72	1.84	0.3352	0.6182	5433	56.1	103
BEL-1_BMa-I	256.4	525.34	2.05	0.2822	0.5782	5603	45.8	94
TREP_CE	1.1	16.76	14.77	0.0012	0.0184	232	4.9	72
INE_WB	27.3	34.33	1.26	0.0300	0.0378	969	28.2	35
MINISAT2_CB	6.1	10.47	1.73	0.0067	0.0115	350	17.3	30
BEL-2_BMa-I	38.1	89.27	2.34	0.0419	0.0983	5789	6.6	15
HELITRON7_CB	7.4	17.78	2.39	0.0082	0.0196	1381	5.4	13
Mariner_HB	7.2	9.32	1.29	0.0080	0.0103	1283	5.6	7
BEL-1_BMa-LTR	1.2	1.89	1.64	0.0013	0.0021	417	2.8	5
BEL-2_ASu-I	3.2	9.36	2.94	0.0035	0.0103	6607	0.5	1
MINISAT4_CB	0.7	0.62	0.92	0.0007	0.0007	500	1.3	1

Supplementary Table 6. Summary of repetitive and low-complexity sequence in the *W. bancrofti* genome.

Category	Overall length (kb)	Adjusted length (kb)	Fold difference between overall length and adjusted length	Percent of genome	Percent of genome (adjusted)
Repetitive elements	1422	5077	3.57	1.75	6.23
Low complexity regions	924	3184	3.45	1.13	3.91

Supplementary Table 7. Known nematode repeats found in the *W. bancrofti* genome.

Repeat	Overall length (kb)	Adjusted length (kb)	Fold difference between overall length and adjusted length	Percent of genome	Percent of genome (adjusted)	Repeat unit length	Estimated #copies	Estimated #copies (adjusted length)
SSPI	1.19	417.66	350.96	0.0015	0.5127	195	6.1	2142
BEL-3_BMa-I	221.79	739.10	3.33	0.2723	0.9073	5433	40.8	136
BEL-1_BMa-LTR	19.41	43.39	2.24	0.0238	0.0533	417	46.5	104
INE_WB	32.43	98.62	3.04	0.0398	0.1211	969	33.5	102
BEL-1_BMa-I	106.11	282.04	2.66	0.1303	0.3462	5603	18.9	50
BEL-3_BMa-LTR	8.27	15.68	1.9	0.0102	0.0192	496	16.7	32
BEL-2_BMa-I	33.97	181.38	5.34	0.0417	0.2227	5789	5.9	31
Mariner_HB	1.80	5.63	3.13	0.0022	0.0069	1283	1.4	4
CeRep52	0.22	0.39	1.74	0.0003	0.0005	224	1.0	2
MBOI	0.05	0.09	1.89	0.0001	0.0001	62	0.8	1

Supplementary Table 11. Summary of repetitive and low-complexity sequence in the *B. malayi* genome.

Category	Overall length (kb)	Percent of genome
Repetitive elements	11337	12.11
Low complexity regions	1054	1.13

Supplementary Table 12. Known nematode repeats found in the *B. malayi* genome.

Repeat	Overall length (kb)	Percent of genome	Repeat unit length	Estimated #copies
MBOI	3140.7	3.354	62	50656
HHAI_BMA	1324.8	1.415	322	4114
BEL-3_BMa-I	582.9	0.622	5433	107
BEL-1_BMa-I	259.9	0.278	5603	46
BEL-2_BMa-I	205.1	0.219	5789	35
BEL-3_BMa-LTR	101.7	0.109	496	205
INE_WB	58.5	0.063	969	60
BEL-1_BMa-LTR	33.9	0.036	417	81
BEL-2_BMa-LTR	30.0	0.032	541	55
TREP_CE	13.5	0.014	232	58
RCS5	1.5	0.002	1428	1

Supplementary Table 15. Nuclear *Wolbachia* transfers (nuwts) in the *L. loa* genome. Their distribution in *B. malayi* (BM), *A. viteae* (AV), and *O. flexuosa* (OF) is also shown.

Name	Class	*L. loa* gene	*wBm* gene	BM	AV	OF
NUWT1	I	LOAG_18970: Cytochome oxidase 1	Wbm0307: Cytochrome oxidase (coxA)	Y	Y	Y
NUWT5	I	LOAG_18725: Succinate dehydroganse	Wbm0448: Succinate dehyodrogenase	Y	Y	N
NUWT6	I	LOAG_03457: NADH dehydrogenase	Wbm0471: NADH dehydrogenase	Y	Y	Y
NUWT7	I	LOAG_03614: Fumurase	Wbm0504: Fumarase	Y	Y	N
NUWT12	I	LOAG_01333: Pyruvate dehyodrgenase	Wbm0666: Pyruvate dehydrogenase	Y	Y	N
NUWT13	I	LOAG_08048: ATP synthase	Wbm0689 F0F1-type ATP synthase	Y	Y	N
NUWT14	I	LOAG_08048: ATP synthase	Wbm0689 F0F1-type ATP synthase	Y	Y	N
NUWT2	II	LOAG_11300: Hypothetical protein	Wbm0309: Membrane fusion	N	N	N
NUWT3	II	LOAG_11300: Hypothetical protein	Wbm0777: Porphobilinogen deaminase	N	N	N
NUWT4	II	LOAG_18709: Hypothetical protein	Wbm0430: DNA-directed RNA polymerase,RpoH	N	N	N
NUWT15	II	LOAG_18709: Hypothetical protein	Wbm0430: DNA-directed RNA polymerase,RpoH	N	N	N
NUWT10	II	LOAG_18265: Hypothetical protein	Wbm0622: NADH:ubiquinone oxidoreductase	N	N	N
NUWT8	III	*L. loa* assembly 293:340-399	no locus tag; folate pseudogene	N	N	N
NUWT9	III	*L. loa* assembly 576:26810-26900	Wbm0552: ATP-dependent protease Clp	N	N	N
NUWT11	III	*L. loa* assembly 316:73448-73505	Wbm0658: Ribosomal protein L35	N	N	N

Supplementary Table 16. Gene statistics of the *L. loa* and *W. bancrofti* genomes. RNA-Seq-related statistics are shown only for *L. loa*. Genes with FPKM (fragments per kilobase of exon per million fragments mapped) >= 1 were considered supported by RNA-Seq.

Category	*L. loa*	*W. bancrofti*
Number of genes	14907	19327[a]
Avg. gene size (bp)	3080	2571
Avg. exon size (bp)	165	167
Avg. intron size (bp)	335	315
Avg. number of exons / gene	6.8	4.5
Avg. gene % GC	33.1	32.7
Avg. exon % GC	39.4	39.4
Avg. intron % GC	29.3	27.7
RNA-Seq supported genes	10468	--
Alternatively spliced genes	502	--
5' UTRs	5238	--
3' UTRs	5123	--
Genes with PFAM, GO, or EC	8627	12700
Genes in functional categories	7714	10839
Secreted proteins	994	781
Number of encoded tRNAs	124	112

[a]due to fragmentation of the genome assembly, the true *W. bancrofti* gene count is estimated to be 14,496–15,075 genes, while the true *L. loa* gene count is estimated to be 14,261 genes.

Supplementary Table 20. Comparison of protein kinases in the *L. loa* and *C. elegans* genomes.

Category	*L. loa*	*C. elegans*
Total number of protein kinases (kinome size)	310	444
Total number of aPKs[a]	19	17
Total number of ePKs[a]	291	427
Number of shared protein kinase orthology groups	218	218
Number of orthologous protein kinases, including inparalogs	250	246
Protein kinases with no ortholog in the other nematode	60	160
Protein kinase with no ortholog, but with related kinase(s) in the other nematode	48	146
Families not present in the other nematode	8	15
Unclassified, potentially novel, kinases	6	10

[a]aPK, atypical protein kinases; ePK, eukaryotic protein kinase superfamiily members

Supplementary Table 24. Nematode phylogenetic profiles of metabolic pathways hypothesized to be involved in the filaria-*Wolbachia* symbiosis.

Gene		*C. elegans*	*C. briggsae*	*P. pacificus*	*M. hapla*	*T. spiralis*	*A. suum*	*B. malayi*	*W. bancrofti*	*L. loa*
Heme biosynthesis										
ALA synthase		–	–	–	–	–	–	–	–	–
Porphobilinogen synthase		–	–	–	–	–	–	–	–	–
Porphobilinogen deaminase		–	–	–	–	–	–	–	–	–
Uroporphyrinogen III synthase		–	–	–	–	–	–	–	–	–
Uroporphyrinogen III decarboxylase		–	–	–	–	–	–	–	–	–
Coproporphyrinogen III oxidase		–	–	–	–	–	–	–	–	–
Protoporphyrinogen oxidase		–	–	–	–	–	–	–	–	–
Ferrochelatase		–	–	–	–	–	–	+[f]	+[f]	+[f]
Flavin biosynthesis (animal)										
Riboflavin kinase	WBGene00011224	+	–	–	+[b]	+	+	+	+	+
FAD synthase	WBGene00011271	+	+[b]	+	+	+	+	+	+	+
Glutathione biosynthesis										
Glutamate-Cysteine ligase	WBGene00001527	+	+	+	+	+	+	+	+	+
Glutathione synthase	WBGene00010941	+	+	+	+	+	+	+	+	+
Purine Synthesis										
Amidophosphoribosyltransferase	WBGene00011407	+	–	–	–	+	+	–[m]	–	–
Phosphoribosylglycinamide formyltransferase	WBGene00018174	+	–[e]	–	+	+	+	–[w]	–[w]	–
Phosphoribosylformylglycinamidine synthase	WBGene00008654	+	–[e]	–	+	+	+	–	–	–
Phosphoribosylaminoimidazole carboxylase	WBGene00015116	+	–	–	+	+	+	–[m]	–	–
Adenylosuccinate lyase[s]	WBGene00011064	+	+	+	+	+	+	+	+	+
IMP cyclohydrolase	WBGene00016957	+	+	+	–	+	+	–	–	–
Purine Interconversion										
Adenylosuccinate lyase[s]	WBGene00011064	+	+	+	+	+	+	+	+	+
IMP dehydrogenase	WBGene00020682	+	+	+	+	+	+	+	+	+
GMP synthase[a]	WBGene00010912	+	+	+	+	+	+	+	+	+
Adenylosuccinate synthase	WBGene00016509	+	+	+	+	+	+	+	+	+
GMP reductase	WBGene00017984	+	+	+	–	+	+	+	+	+
AMP deaminase	WBGene00016415	+	+	+	+	+	+	+	+	+
Purine Salvage										
Adenine phosphoribosyltransferase	WBGene00020557	+	+	+	–	+	+	+	+	+
Hypoxanthine-guanine phosphoribosyltransferase	WBGene00013690	+	–	+	+	+	+	+	+	+[b,r]
Purine-nucleoside phosphorylase	WBGene00019298	+	+	+	+	+	+	+	+	+
Adenosine kinase	WBGene00011128	+	+	+	+	+	–	+	+	+
Adenosine deaminase	WBGene00015551	+	+	–	+	+	+	+[b]	+	+
Pyrimidine Synthesis & Interconversion										
Aspartate carbamoyltransferase	WBGene00004259	+	+	+	–	+	+	–	–	–
Dihydroorotate dehydrogenase	WBGene00020932	+	+	+	–	+	+	+	+	+
Orotidine-5'-phosphate decarboxylase[a]	WBGene00011559	+	+	+	+	+	+	+	+	+

Nucleoside-diphosphatase[a]	WBGene00003254	+	+	+	+	+	+	+	+
Nucleoside-diphosphate kinase[a]	WBGene00009119	+	+	+	+	+	+	+	+
Nucleoside-triphosphatase[a]	WBGene00001823	+	+	+[c]	+	+	+	+	+
GMP/CTP synthase	WBGene00012316	+	+	+	+	+	+	+	+
Cytidylate kinase[a]	WBGene00009531	+	+	+	+	+	+	+	+

[a]multiple genes were predicted to perform the same function in *C. elegans*. In these cases, we selected one of the genes with maximum coverage across nematodes.

[b]identified with tblastn using a cutoff of 1e-10 in the absence of a predicted ortholog

[c]identified with blastp using a cutoff of 1e-10 in the absence of a predicted ortholog

[e]identified with tblastn using a cutoff of 1e-10, but is *E. coli* vector contamination

[f]all filarial nematodes have a laterally transferred ferrochelatase[19]

[m]identified with tblastn using a cutoff of 1e-10, but is mosquito contamination

[r]in addition to tblastn identification, transcription in area supported by low level RNA-Seq

[s]involved in both purine synthesis and interconversion

[w]have a partial insertion of equivalent Wolbachia gene

Supplementary Table 25. *Wolbachia* phylogenetic profiles of metabolic pathways hypothesized to be involved in the filaria-*Wolbachia* symbiosis. Classification of *wBm* genes is consistent with that presented in Foster *et al.*[18].

Gene	*wBm*	*wMel*	*wPip*	*wWb*[a]
Heme biosynthesis				
ALA synthase	Wbm0133	+	+	+
Porphobilinogen synthase	Wbm0373	+	+	–
Porphobilinogen deaminase	Wbm0777	+	+	+
Uroporphyrinogen III synthase	Wbm0728	+	+	+
Uroporphyrinogen III decarboxylase (hemE)	Wbm0001	+	+	+
Coproporphyrinogen III oxidase (hemF)	Wbm0709	+	+	+
Protoporphyrinogen oxidase (hemG)	Wbm0208[b]	+	+	+
Ferrochelatase (hemH)	Wbm0719	+	+	+
Flavin biosynthesis (bacterial)				
3,4-DHBP synthase (ribB)	Wbm0312	+	+	–
GTP cyclohydrolase (ribA)	Wbm0278	+	+	+
pyrimidine deaminase / pyrimidine reductase (ribD)	Wbm0026	+	+	+
riboflavin synthase, beta chain (ribH)	Wbm0189	+	+	–
riboflavin synthase, alpha chain (ribE)	Wbm0083	+	+	–
riboflavin kinase / FMN adenylyltransferase (ribF)	Wbm0416	+	+	+
Glutathione biosynthesis				
Glutamate-Cysteine ligase (gshA)	Wbm0721	+	+	+
Glutathione synthase (gshB)	Wbm0556	+	+	+
Purine biosynthesis & conversion				
Amidophosphoribosyltransferase (purF)	Wbm0255	+	+	+
Phosphoribosylamine-glycine ligase (purD)	Wbm0465	+	+	+
Phosphoribosylglycinamide formyltransferase (purN)	Wbm0420	+	+	–
Phosphoribosylformylglycinamidine synthase (purL)	Wbm0232/0271	+/+	+/+	+/+
Phosphoribosylformylglycinamide cyclo-ligase (purM)	Wbm0226	+	+	+
NCAIR synthase (purK)	Wbm0041	+	+	+
NCAIR mutase (purE)	Wbm0397	+	+	+
SAICAR synthase (purC)	Wbm0227	+	+	+
Adenylosuccinate lyase (purB)	Wbm0503	+	+	+
IMP cyclohydrolase (purH)	Wbm0411	+	+	+
IMP dehydrogenase (guaB)	Wbm0527	+	+	+
GMP synthase (guaA)	Wbm0443	+	+	+
Adenylosuccinate synthase (purA)	Wbm0273	+	+	+
Pyrimidine biosynthesis				
Carbamoyl-phosphate synthase (carA)	Wbm0654	+	+	+
Carbamoyl-phosphate synthase (carB)	Wbm0512	+	+	+
Aspartate carbamoyltransferase (pyrB)	Wbm0385	+	+	+
Dihydroorotase (pyrC)	Wbm0446	+	+	+
Dihydroorotate dehydrogenase (pyrD)	Wbm0098	+	+	+
Orotate phosphoribosyltransferase (pyrE)	Wbm0790	+	+	+
Orotidine-5'-phosphate decarboxylase (pyrF)	Wbm0787	+	+	+
UMP kinase (pyrH)	Wbm0806	+	+	+
Nucleoside-diphosphate kinase[a](ndk)	Wbm0717	+	+	+
GMP/CTP synthase (pyrG)	Wbm0169	+	+	+

[a]as the *wWb* genome was not annotated, genes were identified by searching the *wBm* genes against the *wWb* genome using tblastn with a cutoff of 1e-10
[b]activity present as COG1981[20]

References

1   Klion, A. D., Raghavan, N., Brindley, P. J. & Nutman, T. B. Cloning and characterization of a species-specific repetitive DNA sequence from *Loa loa*. *Mol. Biochem. Parasitol.* **45**, 297-305 (1991).

2   Zhong, M. *et al.* A polymerase chain reaction assay for detection of the parasite *Wuchereria bancrofti* in human blood samples. *Am. J. Trop. Med. Hyg.* **54**, 357-363 (1996).

3   Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**, 1061-1067 (2007).

4   Li, L. *et al.* Possible cross-species transmission of circoviruses and cycloviruses among farm animals. *J. Gen. Virol.* **92**, 768-772 (2011).

5   Ghedin, E. *et al.* Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* **317**, 1756-1760 (2007).

6   Villa, F. *et al.* Structural insights into the recognition of substrates and activators by the OSR1 kinase. *EMBO Rep.* **8**, 839-845 (2007).

7   Sheldon, A. L., Zhang, J., Fei, H. & Levitan, I. B. SLOB, a SLOWPOKE channel binding protein, regulates insulin pathway signaling and metabolism in *Drosophila*. *PLoS One* **6**, e23343 (2011).

8   Kasten, M. & Giordano, A. Cdk10, a Cdc2-related kinase, associates with the Ets2 transcription factor and modulates its transactivation activity. *Oncogene* **20**, 1832-1838 (2001).

9   Sese, M. *et al.* The Cdi/TESK1 kinase is required for Sevenless signaling and epithelial organization in the *Drosophila* eye. *J. Cell Sci.* **119**, 5047-5056 (2006).

10   Adachi, T. *et al.* Reversal of salt preference is directed by the insulin/PI3K and Gq/PKC signaling in *Caenorhabditis elegans. Genetics* **186**, 1309-1319 (2010).

11   Ortiz, C. O. *et al.* Lateralized gustatory behavior of *C. elegans* is controlled by specific receptor-type guanylyl cyclases. *Curr. Biol.* **19**, 996-1004 (2009).

12   Jiu, Y. M. *et al.* Insulin-like signaling pathway functions in integrative response to an olfactory and a gustatory stimuli in *Caenorhabditis elegans. Protein Cell* **1**, 75-81 (2010).

13   Chou, C. M. *et al.* Expression and characterization of a brain-specific protein kinase BSK146 from zebrafish. *Biochem. Biophys. Res. Commun.* **340**, 767-775 (2006).

14   Lu, X. & Li, Y. *Drosophila* Src42A is a negative regulator of RTK signaling. *Dev. Biol.* **208**, 233-243 (1999).

15   van der Linden, A. M. *et al.* The EGL-4 PKG acts with KIN-29 salt-inducible kinase and protein kinase A to regulate chemoreceptor gene expression and sensory behaviors in *Caenorhabditis elegans. Genetics* **180**, 1475-1491 (2008).

16   Meier, B. & Ahmed, S. Checkpoints: chromosome pairing takes an unexpected twist. *Curr. Biol.* **11**, R865-868 (2001).

17   Gilliland, W. D. *et al.* The multiple roles of mps1 in *Drosophila* female meiosis. *PLoS Genet.* **3**, e113 (2007).

18   Foster, J. *et al.* The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.* **3**, e121 (2005).

19   Slatko, B. E., Taylor, M. J. & Foster, J. M. The *Wolbachia* endosymbiont as an anti-filarial nematode target. *Symbiosis* **51**, 55-65 (2010).

20   Boynton, T. O. *et al.* Discovery of a gene involved in a third bacterial protoporphyrinogen oxidase activity through comparative genomic analysis and functional complementation. *Appl. Environ. Microbiol.* **77**, 4795-4801 (2011).

21   Bourbouloux, A., Shahi, P., Chakladar, A., Delrot, S. & Bachhawat, A. K. Hgt1p, a high affinity glutathione transporter from the yeast *Saccharomyces cerevisiae. J. Biol. Chem.* **275**, 13259-13265 (2000).

22   Jex, A. R. *et al. Ascaris suum* draft genome. *Nature* **479**, 529-533 (2011).

23   Rajagopal, A. *et al.* Haem homeostasis is regulated by the conserved and concerted functions of HRG-1 proteins. *Nature* **453**, 1127-1131 (2008).

24   Severance, S. *et al.* Genome-wide analysis reveals novel genes essential for heme homeostasis in *Caenorhabditis elegans. PLoS Genet.* **6**, e1001044 (2010).

25   Stolz, J. & Vielreicher, M. Tpn1p, the plasma membrane vitamin B6 transporter of *Saccharomyces cerevisiae. J. Biol. Chem.* **278**, 18990-18996 (2003).

26   Rodionov, D. A. *et al.* A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.* **191**, 42-51 (2009).

27   Wu, M. *et al.* Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2**, E69 (2004).

28   McNulty, S. N. *et al.* Endosymbiont DNA in endobacteria-free filarial nematodes indicates ancient horizontal genetic transfer. *PLoS One* **5**, e11029 (2010).

29   Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755-763 (1998).

30   Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641-1650 (2009).

31   Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440-9445 (2003).

32   Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797 (2004).